Human WARP

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 March 2001 (15.03.2001)

PCT

(10) International Publication Number
WO 01/18022 A1

(51) International Patent Classification⁷: C07H 21/04, 21/02, C07K 5/00, 14/00, C12Q 1/68, C12N 1/12, 15/63, 15/85, 15/86

(21) International Application Number: PCT/US00/24008

(22) International Filing Date: 31 August 2000 (31.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/152,317 3 September 1999 (03.09.1999) US
60/152,315 3 September 1999 (03.09.1999) US

(71) Applicant *(for all designated States except US)*: **HUMAN GENOME SCIENCES, INC.** [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **NI, Jian** [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). **BAKER, Kevin, P.** [GB/US]; 14006 Indian Run Drive, Darnestown, MD 20878 (US). **BIRSE, Charles, E.** [GB/US]; 13822 Saddleview Drive, North Potomac, MD 20878 (US). **FISCELLA, Michele** [US/US]; 6308 Redwing Road, Bethesda, MD 20817 (US). **KOMATSOULIS, George, A.** [US/US]; 9518 Garwood Street, Silver Spring, MD 20901 (US). **ROSEN, Craig, A.** [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). **SOPPET, Daniel, R.** [US/US]; 15050 Stillfield Place, Centreville, MD 22020 (US). **YOUNG, Paul, E.** [US/US]; 122 Beckwith Street, Gaithersburg, MD 20878 (US). **EBNER, Reinhard** [DE/US]; 9906 Shelburne Terrace, #316, Gaithersburg, MD 20878 (US). **DUAN, D., Roxanne** [US/US]; 5515 Northfield Road, Bethesda, MD 20817 (US). **OLSEN, Henrik, S.** [DK/US]; 182 Kendrick Place, #24, Gaithersburg, MD 20878 (US). **LAFLEUR, David, W.** [US/US]; 3142 Quesada Street, N.W., Washington, DC 20015 (US). **MOORE, Paul, A.** [GB/US]; 19005 Leatherbark Drive, Germantown, MD 20874 (US). **SHI, Yanggu** [US/US]; Apt. 102, 437 West Side Drive, Gaithersburg, MD 20878 (US). **WEI, Ying-Fei** [CN/US]; 242 Gravatt Drive, Berkeley, CA 94705 (US). **FLORENCE, Kimberly, A.** [US/US]; 12805 Atlantic Avenue, Rockville, MD 20851 (US).

(74) Agents: **HOOVER, Kenley, K.** et al.; c/o Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*
— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*
— *With (an) indication(s) in relation to deposited biological material furnished under Rule 13bis separately from the description.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/18022 A1

(54) Title: 52 HUMAN SECRETED PROTEINS

(57) Abstract: The present invention relates to novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating diseases, disorders, and/or conditions related to these novel human secreted proteins.

52 Human Secreted Proteins

Field of the Invention

This invention relates to newly identified polynucleotides, polypeptides encoded by these polynucleotides, antibodies that bind these polypeptides, uses of such polynucleotides, polypeptides, and antibodies, and their production.

Background of the Invention

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoeitin. Thus, in light of

the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical diseases, disorders, and/or conditions by using secreted proteins or the genes that encode them.

Summary of the Invention

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant and synthetic methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting diseases, disorders, and/or conditions related to the polypeptides and polynucleotides, and therapeutic methods for treating such diseases, disorders, and/or conditions. The invention further relates to screening methods for identifying binding partners of the polypeptides.

Detailed Description

Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide. The term "isolated" does not refer to genomic or cDNA libraries, whole cell total or mRNA preparations, genomic DNA preparations (including those separated by electrophoresis and transferred onto blots), sheared whole cell genomic DNA preparations or other compositions where the art demonstrates no distinguishing features of the polynucleotide/sequences of the present invention.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

In specific embodiments, the polynucleotides of the invention are at least 15, at least 30, at least 50, at least 100, at least 125, at least 500, or at least 1000 continuous nucleotides but are less than or equal to 300 kb, 200 kb, 100 kb, 50 kb, 15 kb, 10 kb, 7.5 kb, 5 kb, 2.5 kb, 2.0 kb, or 1 kb, in length. In a further embodiment, polynucleotides of the invention comprise a portion of the coding sequences, as disclosed herein, but do not comprise all or a portion of any intron. In another embodiment, the polynucleotides comprising coding sequences do not contain coding sequences of a genomic flanking gene (i.e., 5' or 3' to the gene of interest in the genome). In other embodiments, the polynucleotides of the invention do not contain the coding sequence of more than 1000, 500, 250, 100, 50, 25, 20, 15, 10, 5, 4, 3, 2, or 1 genomic flanking gene(s).

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As

shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard, Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42 degree C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 μg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65 degree C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37 degree C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH_2PO_4; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 ug/ml salmon sperm blocking DNA; followed by washes at 50 degree C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone generated using oligo dT as a primer).

The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched , for example, as a result of

ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

Many proteins (and translated DNA sequences) contain regions where the amino acid composition is highly biased toward a small subset of the available

residues. For example, membrane spanning domains and signal peptides (which are also membrane spanning) typically contain long stretches where Leucine (L), Valine (V), Alanine (A), and Isoleucine (I) predominate. Poly-Adenosine tracts (polyA) at the end of cDNAs appear in forward translations as poly-Lysine (poly-K) and poly-Phenylalanine (poly-F) when the reverse complement is translated. These regions are often referred to as "low complexity" regions.

Such regions can cause database similarity search programs such as BLAST to find high-scoring sequence matches that do not imply true homology. The problem is exacerbated by the fact that most weight matrices (used to score the alignments generated by BLAST) give a match between any of a group of hydrophobic amino acids (L,V and I) that are commonly found in certain low complexity regions almost as high a score as for exact matches.

In order to compensate for this, BLASTX.2 (version 2.0a5MP-WashU) employs two filters ("seg" and "xnu") which "mask" the low complexity regions in a particular sequence. These filters parse the sequence for such regions, and create a new sequence in which the amino acids in the low complexity region have been replaced with the character "X". This is then used as the input sequence (sometimes referred to herein as "Query" and/or "Q") to the BLASTX program. While this regime helps to ensure that high-scoring matches represent true homology, there is a negative consequence in that the BLASTX program uses the query sequence that has been masked by the filters to draw alignments.

Thus, a stretch of "X"s in an alignment shown in the following application does not necessarily indicate that either the underlying DNA sequence or the translated protein sequence is unknown or uncertain. Nor is the presence of such stretches meant to indicate that the sequence is identical or not identical to the sequence disclosed in the alignment of the present invention. Such stretches may simply indicate that the BLASTX program masked amino acids in that region due to the detection of a low complexity region, as defined above. In all cases, the reference sequence(s) (sometimes referred to herein as "Subject", "Sbjct", and/or "S") indicated in the specification, sequence table (Table 1), and/or the deposited clone is (are) the definitive embodiment(s) of the present invention, and should not be construed as

limiting the present invention to the partial sequence shown in an alignment, unless specifically noted otherwise herein.

Polynucleotides and Polypeptides of the Invention

FEATURES OF PROTEIN ENCODED BY GENE NO: 1

The translation product of this gene shares sequence homology with env protein (see, e.g., Genbank accession number AAD34324.1 (AF108843); all references available through this accession are hereby incorporated by reference herein.), a protein with similarity to retroviral envelope glycoproteins.

The polypeptide of this gene has been determined to have a transmembrane domain at about amino acid position 493 to about 509 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing from about amino acids 510 to about 563 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type Ia membrane proteins.

This gene is expressed primarily in fetal tissues, placenta, fetal liver spleen, infant brain, and total fetus and to a lesser extent in tumors (poorly differentiated ovarian adenocarcinoma and endometrial tumor), human adult (K.Okubo) and PC3 prostate cell line.

Polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include but are not limited to: fetal development disorders, cancer and other proliferative disorders, particularly endometrial and ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endometrium and ovary, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell

types (e.g., fetal, reproductive, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise, or alternatively consist of, one or more immunogenic epitopes shown in SEQ ID NO: 83 as residues: Gln-88 to Lys-97, Glu-128 to Ser-133, Asn-166 to Pro-175, Thr-191 to Asn-196, Asn-207 to Lys-212, Cys-232 to Gly-238, Ala-256 to Ala-263, Thr-268 to Thr-280, Pro-311 to Cys-317, Val-347 to Leu-362, Glu-396 to Leu-406, Pro-429 to Ala-436, Ala-464 to Lys-469, Arg-513 to Asn-520. Polynucleotides encoding said polypeptides are also encompassed by the invention.

The tissue distribution and homology to retroviral envelope proteins indicates that polynucleotides and polypeptides corresponding to this gene would be useful for diagnosis, detection, prevention and/or treatment of cancer and other proliferative disorders, particularly of the endometrium and ovary.

The tissue distribution in infant brain indicates the protein product of this clone would be useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival.

The expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and would be useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:11 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence

would be cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2205 of SEQ ID NO:11, b is an integer of 15 to 2219, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:11, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 2

This gene shares sequence homology with members of the B7 family of ligands (i.e., B7-1 (See Genbank Accession 507873)). These proteins and their corresponding receptors play vital roles in the growth, differentiation and death of T cells. For example, some members of this family (i.e., B7-H1) are involved in co-stimulation of the T cell response, as well as inducing increased cytokine production. Therefore, antagonists such as antibodies or small molecules directed against the translation product of this gene are useful for treating T cell mediated immune system disorders.

In additional nonexclusive embodiments, polypeptides of the invention comprise, or alternatively consist of, the following amino acid sequence: LEVQVPEDPVVALVGTDATLCCSFSPEPGFSLAQLNLIWQLTDTKQLVHSFAE GQDQGSAYANRTALFPDLLAQGNASLRLQRVRVADEGSFTCFVSIRDFGSAA VSLQVAAPYSKPSMTLEPNKDLRPGDTVTITCSSYQGYPEAEVFWQDGQGVP LTGNVTTSQMANEQGLFDVHSILRVVLGANGTYSCLVRNPVLQQDAHSSVTI TGQPMTF (SEQ ID NO: 158). Moreover, fragments and variants of these polypeptides (such as, for example, fragments as described herein, polypeptides at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to these polypeptides and polypeptides encoded by the polynucleotide which hybridizes, under stringent conditions, to the polynucleotide encoding these polypeptides , or the complement there of are encompassed by the invention. Antibodies that bind polypeptides of the

invention are also encompassed by the invention. Polynucleotides encoding these polypeptides are also encompassed by the invention.

Also preferred are polypeptides comprising, or alternatively consisting of, fragments of the mature extracellular portion of the protein demonstrating functional activity. Polynucleotides encoding these polypeptides are also encompassed by the invention.

Such functional activities include, but are not limited to, biological activity (e.g., T cell costimulatory activity, ability to bind ICOS, and ability to induce or inhibit cytokine production), antigenicity, immunogenicity (ability to generate antibody which binds to a polypeptide of the invention), ability to form multimers with polypeptides of the invention, and ability to bind to a receptor or ligand for a polypeptide of the invention.

Additionally, the translation product of this gene shares sequence homology with butyrophilin and butyrophilin-like molecules (See, e.g., Genbank Accession No. emb|CAB38473.1| (AL034394) dJ1077I5.1 and gb|AAC05288.1| (AF050157); in addition to the following Geneseq Accession Nos. W46488, W97816, W71592, and W78917; all information and references available through these accessions are hereby incorporated herein by reference):

```
gb|AAC05288.1| (AF050157) butyrophilin-like (Mus musculus) >sp|O70355|O70355
            BUTYROPHILIN-LIKE (FRAGMENT).
            Length = 452

  Plus Strand HSPs:

 Score = 255 (89.8 bits), Expect = 2.9e-23, Sum P(2) = 2.9e-23
 Identities = 80/292 (27%), Positives = 137/292 (46%), Frame = +1

Query:   613 GPGDMVTITCSSYQGYPEAEVFWQDGQGVPLTGNVTTSQMANEQGLFDVHSILRVVLGAN 792
            G G+ V + C+S   +PE EV W+   G  L  + + + +  E GLF V   L V   +
Sbjct:   156 GEGE-VQLVCTSRGWFPEPEVHWEGIWGEKLM-SFSENHVPGEDGLFYVEDTLMVRNDSV 213

Query:   793 GTYSCLVRNPVLQQDAHSSVTITPQ-RSPTGAVEVQVPEDPVVALVGTDATLHCSFSPEP 969
             T SC + +  L++   +++ + + ++   +V V     P    VG +  L C  SP+
Sbjct:   214 ETISCFIYSHGLRETQEATIALSERLQTELASVSVIGHSQPSPVQVGENIELTCHLSPQT 273

Query:   970 GFSLTQLNLIWQLTDTKQLVHSFTEGR----DQGSAYANRTALFPDLLAQGNASLRLQRV 1137
               L + W  +     VH +  G      +Q   Y  RT+L  D + +G  +L++
Sbjct:   274 --DAQNLEVRWLRSRYYPAVHVYANGTHVAGEQMVEYKGRTSLVTDAIHEGKLTLQIHNA 331

Query:  1138 RVADEGSFTCFVSIRD--FGSAAVSLQVAAPYSKPSMTLEPNKDLRPGDTVTITCSSYRG 1311
            R +DEG + C    +D  +  A V +QV A  S P +T E  KD   G  + + C+S
Sbjct:   332 RTSDEGQYRCLFG-KDGVYQEARVDVQVMAVGSTPRITREVLKD---GG-MQLRCTSDGW 386

Query:  1312 YPEAEVFWQDGQGVPLTGNVTTSQMANEQGLFDVHSVLRVVLGANGTYSCLVRNPVLQQ 1488
            +P  V W+D  G  +       Q  +++ LF V  +L V  G+     +C +  P+ Q+
Sbjct:   387 FPRPHVQWRDRDGKTMPSFSEAFQQGSQE-LFQVETLLLVTNGSMVNVTCSISLPLGQE 444
```

```
 Score = 194 (68.3 bits), Expect = 4.6e-11, P = 4.6e-11
 Identities = 58/210 (27%), Positives = 103/210 (49%), Frame = +1

Query:   901 PEDPVVALVGTDATLHCSFSPEPGFSLTQLNLIWQLTDTKQLVHSFTEGRD-QG---SAY 1068
             P  P++A VG DA L C    P+    +  + + W  +D    V  + +G +  G     Y
Sbjct:    34 PNLPILAKVGEDALLTCQLLPKR--TTAHMEVRWYRSDPDMPVIMYRDGAEVTGLPMEGY 91

Query:  1069 ANRTALFPDLLAQGNASLRLQRVRVADEGSFTCFVSIRDFG-SAAVSLQVAAPYSKPSMT 1245
               R     D   +G+  +L++++V+ +D+G + C     D+    +V LQVAA  S P++
Sbjct:    92 GGRAEWMEDSTEEGSVALKIRQVQPSDDGQYWCRFQEGDYWRETSVLLQVAALGSSPNIH 151

Query:  1246 LEPNKDLRPGDTVTITCSSYRGYPEAEVFWQDGQGVPLTGNVTTSQMANEQGLFDVHSVL 1425
             +E    L  G+ V + C+S    +PE EV W+   G  L  + + + +  E GLF V   L
Sbjct:    152 VE---GLGEGE-VQLVCTSRGWFPEPEVHWEGIWGEKLM-SFSENHVPGEDGLFYVEDTL 206

Query:  1426 RVVLGANGTYSCLVRNPVLQQDAHGSVTITGQPMT 1530
              V   +  T SC + +  L++     ++ ++ +  T
Sbjct:    207 MVRNDSVETISCFIYSHGLRETQEATIALSERLQT 241


 Score = 105 (37.0 bits), Expect = 0.24, P = 0.21
 Identities = 30/100 (30%), Positives = 44/100 (44%), Frame = +2

Query:   254 PVVALVGTDATLCCSFSPEPGFSLAQLNLIWQLTDTKQLVHSFAEGQ----DQGSAYANR 421
             P    VG +  L C  SP+       L + W  +     VH +A G      +Q    Y  R
Sbjct:    254 PSPVQVGENIELTCHLSPQT--DAQNLEVRWLRSRYYPAVHVYANGTHVAGEQMVEYKGR 311

Query:   422 TALFLDLLAQGNASLRLQSVRVADEGQLHLLREHPGFRQRCR 547
             T+L  D + +G   +L++ + R +DEGQ   L    G  Q  R
Sbjct:    312 TSLVTDAIHEGKLTLQIHNARTSDEGQYRCLPGKDGVYQEAR 353


 Score = 97 (34.1 bits), Expect = 2.9e-23, Sum P(2) = 2.9e-23
 Identities = 25/88 (28%), Positives = 44/88 (50%), Frame = +2

Query:   245 PEDPVVALVGTDATLCCSFSPEPGFSLAQLNLIWQLTDTKQLVHSFAEGQD-QG---SAY 412
             P  P++A VG DA L C    P+    + A + + W  +D    V  + +G +  G     Y
Sbjct:    34 PNLPILAKVGEDALLTCQLLPKR--TTAHMEVRWYRSDPDMPVIMYRDGAEVTGLPMEGY 91

Query:   413 ANRTALFLDLLAQGNASLRLQSVRVADEGQ 502
               R     D   +G+  +L+++ V+ +D+GQ
Sbjct:    92 GGRAEWMEDSTEEGSVALKIRQVQPSDDGQ 121
```

Butyrophilin is thought to be important in the process of lactation and milk secretion. Based on the sequence similarity, the translation product of this clone is expected to share at least some biological activities with butyrophilin and/or oligodendrite proteins. Such activities are known in the art, some of which are described elsewhere herein.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise, or alternatively consist of, the following amino acid sequence:

ARLGRVPESQSRRGAAGAAFHHGEPSCQPPHRKMLRRRGSPGMGVHVGAAL GALW

FCLTGALEVQVPEDPVVALVGTDATLCCSFSPEPGFSLAQLNLIWQLTDTKQL

VHSFAEGQDQGSAYANRTALFLDLLAQGNASLRLQSVRVADEGQLHLLREH PGFRQRCRQPAGGRSLLEAQHDPGAQQGPAARGTW (SEQ ID NO: 155). Polynucleotides encoding these polypeptides are also encompassed by the invention.

In specific embodiments, polypeptides of the invention comprise, or alternatively consist of, the following amino acid sequence: PWSPTRTCGPGDMVTITCSSYQGYPEAEVFWQDGQGVPLTGNVTTSQMANE QGLFDVHSILRVVLGANGTYSCLVRNPVLQQDAHSSVTITPQRSPTGAVEVQ VPEDPVVALVGTDATLHCSFSPEPGFSLTQLNLIWQLTDTKQLVHSFTEGRDQ GSAYANRTALFPDLLAQGNASLRLQRVRVADEGSFTCFVSIRDFGSAAVSLQ VAAPYSKPSMTLEPNKDLRPGDTVTITCSSYRGYPEAEVFWQDGQGVPLTGN VTTSQMANEQGLFDVHSVLRVVLGANGTYSCLVR NPVLQQDAHGSVTITGQPMTFPPEALWVTVGLSVCLIALLVALPFVCWRKIK QSCEEENAGAEDQDGEGE GSKTALQPLKHSDSKEDDGQEIA (SEQ ID NO: 156). Moreover, fragments and variants of these polypeptides (such as, for example, fragments as described herein, polypeptides at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to these polypeptides and polypeptides encoded by the polynucleotide which hybridizes, under stringent conditions, to the polynucleotide encoding these polypeptides , or the complement there of are encompassed by the invention. Antibodies that bind polypeptides of the invention are also encompassed by the invention. Polynucleotides encoding these polypeptides are also encompassed by the invention.

The gene encoding the disclosed cDNA is believed to reside on chromosome 15. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 15.

This gene is expressed primarily in dendritic cells and to a lesser extent in fetal liver and spleen, normal colon, and normal liver. It is also expressed in various tumors including ovary, glioblastoma, germ cell tumors, pancreatic tumor, and germinal center B-cell cancer.

Polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include but are not limited to cancer and immune disorders including autoimmune diseases and immuno-deficiency

disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise, or alternatively consist of, one or more immunogenic epitopes shown in SEQ ID NO: 84 as residues: Glu-72 to Gly-77, Arg-115 to Arg-125, His-138 to Pro-146. Polynucleotides encoding said polypeptides are also encompassed by the invention.

The dendritic cell distribution and homology to the butyrophilin family indicates that polynucleotides and polypeptides corresponding to this gene are useful for down-regulation or stimulation of the immune-response. Dendritic cells play a pivotal role in immune surveillance- they are responsible for the capture and processing of antigens from the periphery and subsequent presentation of these antigens to B and T lymphocytes in lymphoid organs. Dendritic cells also produce and secrete numerous immuno-modulatory proteins. The butyrophilin family appears to have a receptor like structure having an extracellular domain, transmembrane domain and intracellular region. The encoded protein may act as a membrane bound receptor to mediate the interaction of dendritic cells with other cells of the immune system. This interaction could be with either soluble factors produced by other immune cells or with membrane proteins present on other immune cells. Such interactions may result in a stimulation or down-regulation of dendritic cell function. Subsequently the immune system may be stimulated to respond against specific antigens, or the response may dampened as is seen in tolerance of self- antigens. The inability to effectively inhibit immune responses to self antigens could result in auto-immune disease. Conversely the inability to stimulate correct responses could result in an immuno-deficiency syndrome and subsequent susceptibility to infectious agents.

Additionally, the expression of this gene in numerous tumors may reflect the role that this molecule plays in the body's normal anti-tumor surveillance system; tumor cells may express this protein in order to stimulate an immune response (e.g.; targeting of cytotoxic T-cells against the tumor cells). Alternately, the molecule may be used by tumors to dampen the cytotoxic immune response and thus be a means by which tumors escape killing.

Moreover, the tissue distribution in fetal liver spleen and germinal center B-cell indicates the protein product of this clone is useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "Infectious Disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional

supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:12 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence would be cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3422 of SEQ ID NO:12, b is an integer of 15 to 3436, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:12, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 3

The translation product of this gene shares sequence homology with matrilin and other cartilage matrix proteins (see, e.g, Genbank Accession Nos. emb|CAA06889.1| (AJ006140); and/or emb|CAA30915.1|; all references available through these accessions are hereby incorporated in their entirety by reference herein). Matrilins are members of a superfamily with von Willebrand factor type A-like modules, which is thought to be important in forming an extracellular, filamentous network.

Moreover, the translation product of this gene also shares sequence homology with the kidney injury associated molecule (KIM) protein (See Geneseq Accession No. W86326; all references and information available through this accession are hereby incorporated herein by reference). Based on the sequence similarity, the translation product of this clone is expected to share at least some biological activities with matrilin, cartilage matrix proteins and KIM proteins. Such activities are known in the art, some of which are described elsewhere herein.

In specific embodiments, polypeptides of the invention comprise, or alternatively consist of, an amino acid sequence selected from the group:

KXPCXYRSGIPGSTHASVPSAPRPSRAMLPWTAXGLALSLRLALARSGAERG
PPASAPRGDLMFLLDSSASVSHYEFSRVREFVGQLVAPLPLGTGALRASLVHV
GSRPYTEFPFGQHSSGEAAQDAVRASAQRMGDTHTGLALVYAKEQLFAEAS
GARPGVPKVLVWVTDGGSSDPVGPPMQELKDLGVTVFIVSTGRGNFLELSAA
ASAPAEKHLHFVDVDDLHIIVQELRGSILDAMRP (SEQ ID NO: 159);
APAWGGPQGRWSRHLSPTPALWAPLAGHLMLQQTAVPWHRPAPGQCGCHP
CAGQKHAPHPGQPHPSCAGRRGTRCMADCPRAPDWHAGPRCPGAVEPPAAP
QTPEPGRTRSERRWLSCPAGTSGPLGGLMLVDRAPRRSAPAPAASSGPGRXPS
RGASRARDGARSARTRGSTREFRTGXCRVXSX (SEQ ID NO: 160),
HASVPSAPRPSRAMLPWTALGLALSLRLALARSGAERGPPASAPRGDLMFLL
DSSASVSHYEFSRVREFVGQLVAPLPLGTGALRASLVHVGSRPYTEFPFGQHS
SGEAAQDAVRASAQRMGDTHTGLALVYAKEQLFAEASGARPGVPKVLVWV
TDGGSSDPVGPPMQELKDLGVTVFIVSTGRGNFLELSAAASAPAEKHLHFVD
VDDLHIIVQELRGSILDAM (SEQ ID NO: 165); FLLDSSASVSHYEFSRVR (SEQ ID NO: 161), GALRASLVHVGSRP (SEQ ID NO: 162), GVPKVLVWVTDG (SEQ ID NO: 163), and VGPPMQELKDLGVT (SEQ ID NO: 164). Moreover, fragments and variants of these polypeptides (such as, for example, fragments as described herein, polypeptides at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to these polypeptides, or polypeptides encoded by a polynucleotide which hybridizes, under stringent conditions, to the polynucleotide encoding these polypeptides) are encompassed by the invention. Antibodies that bind polypeptides of the invention and polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in uterus, brain, lung, colon, kidney, placenta, dendritic cells.

Polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include but are not limited to: renal, neural, endothelial, developmental, and reproductive diseases and/or disorders, particularly disorders resulting from tissue structural damages or abnormalities, Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the uterus, placenta, kidney, lung, brain, and colon, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., renal, neural, endothelial, developmental, reproductive, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution kidney, combined with the homology to the matrilin and KIM proteins indicates that polynucleotides and polypeptides corresponding to this gene would be useful for treatment, prevention, detection and/or diagnosis of disorders involving tissues with structural damages or abnormalities, particularly organs or tissues such as uterus, placenta, kidney, lung, brain, and colon. Matrilin may be also involved in extracellular transport, storage, barrier of molecular factors such as growth factors, hormones, thereby modulating the organ functions. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", "Infectious Disease", and "Regeneration" sections below, in Example 11, 19, and 20, and elsewhere herein.

In addition expression in the placenta indicates that polynucleotides and/or polypeptides corresponding to this gene would be useful in treating, preventing, detecting and/or diagnosing placental related function or diseases, e.g. induced abortion or spontaneous abortion; hyperplastic abnormalities; factors involved in circulation, nutrient transport; prevention of multiple gestation; gestational trophoblastic diseases, such as hydatidiform mole as well as placental site trophoblastic tumor and chriocarcinoma; uterus related function, e.g., disorders during the menstrual cycle or pregnancy, inflammatory changes, such as pyometra, endometritis and dysfunctional bleeding; contraceptives, abortion and birth control; infertility caused by blastocyst, embryo or fetus implantation problems; utilities in surrogate pregnancy; tumors or hyperplasia of the uterus, with epithelium, stroma or smooth muscle origins; brain related functions, e.g., trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, toxic neuropathies induced by neurotoxins, inflammatory diseases such

as meningitis and encephalitis, demyelinating diseases, neurodegenerative diseases such as Parkinson's disease, Huntington's disease, Alzheimer's disease, peripheral neuropathies, multiple sclerosis, neoplasia of neuroectodermal origin, etc; as well as diseases implicated in lung, colon functions. Polynucleotides and/or polypeptides of the invention can be used to promote growth and/or survival of damaged tissue (e.g., renal tissue), since KIM proteins are upregulated in injured or regenerating (especially renal) tissues. Fusion proteins of the invention, conjugates, antibodies and vectors can also be used therapeutically, e.g., these or KIM proteins (or a protein having KIM activity) may be included with an acceptable carrier in pharmaceutical compositions, useful for therapy/prophylaxis of conditions associated with dysfunction/dysregulation of genes or proteins of the invention, especially renal diseases or impairments of renal function in humans (e.g., acute renal failure, acute nephritis). The polynucleotides can be used to produce antisense sequences which, when internalized into cells, can disrupt expression of a cellular gene, also useful in therapy (e.g., to block the growth of tumors dependent on polynucleotides or polypeptides of the invention for growth) or compositions. The proteins and polynucleotides would be useful diagnostically e.g., to detect and quantify renal injury/disease (indicative of increased risk, or presence of, renal injury or impaired function), or abnormal responses to tissue injury (indicative of increased risk, or presence of, an autoimmune response or abnormal tissue growth arising from/affecting renal tissue). The proteins can also be used to locate cells producing the invention (especially specific loci, e.g., tissue masses abnormally producing/expressing polynucleotide or polypeptides of the invention such as tumors arising from/affecting renal tissue), by contacting cells with an imaginable reagent which binds to polynucleotides or polypeptides of the invention and imaging reagent accumulation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

related to SEQ ID NO:13 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence would be cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 720 of SEQ ID NO:13, b is an integer of 15 to 734, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:13, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 4

The translation product of this gene shares sequence homology with Liv-1 which is thought to be an estrogen-regulated gene associated with breast cancer. The polypeptide of this gene has been determined to have seven transmembrane domains at about amino acid positions 3-19, 400-436, 433-457, 493-512, 736-753, 758-781, and/or 800-827 of the amino acid sequence referenced in Table 1 for this gene. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type IIIa membrane proteins.

Included in this invention as preferred domains are zinc finger, C2H2 type, and cytochrome c family heme-binding site signature domains, which were identified using the ProSite analysis tool (Copyright, Swiss Institute of Bioinformatics). 'Zinc finger' domains [1-5] are nucleic acid-binding protein structures first identified in the Xenopus transcription factor TFIIIA. These domains have since been found in numerous nucleic acid-binding proteins.

A zinc finger domain is composed of 25 to 30 amino-acid residues. There are two cysteine or histidine residues at both extremities of the domain, which are involved in the tetrahedral coordination of a zinc atom. It has been proposed that such a domain interacts with about five nucleotides.

A schematic representation of a zinc finger domain is shown below:

x x x x x x x x x x x x C H x / x x Zn x x / x C H x x x x x x x x x x

Gln Leu Val His Ser Phe Ala Glu Gly Gln Asp Gln Gly Ser Ala Tyr
65 70 75 80

Ala Asn Arg Thr Ala Leu Phe Leu Asp Leu Leu Ala Gln Gly Asn Ala
85 90 95

Ser Leu Arg Leu Gln Ser Val Arg Val Ala Asp Glu Gly Gln Leu His
100 105 110

Leu Leu Arg Glu His Pro Gly Phe Arg Gln Arg Cys Arg Gln Pro Ala
115 120 125

Gly Gly Arg Ser Leu Leu Glu Ala Gln His Asp Pro Gly Ala Gln Gln
130 135 140

Gly Pro Ala Ala Arg Gly Thr Trp
145 150

<210> 85
<211> 215
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (7)
<223> Xaa equals any of the naturally occurring L-amino acids

<400> 85
Met Leu Pro Trp Thr Ala Xaa Gly Leu Ala Leu Ser Leu Arg Leu Ala
1 5 10 15

Leu Ala Arg Ser Gly Ala Glu Arg Gly Pro Pro Ala Ser Ala Pro Arg
20 25 30

Gly Asp Leu Met Phe Leu Leu Asp Ser Ser Ala Ser Val Ser His Tyr
35 40 45

Glu Phe Ser Arg Val Arg Glu Phe Val Gly Gln Leu Val Ala Pro Leu
50 55 60

Pro Leu Gly Thr Gly Ala Leu Arg Ala Ser Leu Val His Val Gly Ser
65 70 75 80

Arg Pro Tyr Thr Glu Phe Pro Phe Gly Gln His Ser Ser Gly Glu Ala
85 90 95

Ala Gln Asp Ala Val Arg Ala Ser Ala Gln Arg Met Gly Asp Thr His
100 105 110

Thr Gly Leu Ala Leu Val Tyr Ala Lys Glu Gln Leu Phe Ala Glu Ala
115 120 125

Ser Gly Ala Arg Pro Gly Val Pro Lys Val Leu Val Trp Val Thr Asp
130 135 140

Gly Gly Ser Ser Asp Pro Val Gly Pro Pro Met Gln Glu Leu Lys Asp
145 150 155 160

Leu Gly Val Thr Val Phe Ile Val Ser Thr Gly Arg Gly Asn Phe Leu
165 170 175

Glu Leu Ser Ala Ala Ala Ser Ala Pro Ala Glu Lys His Leu His Phe
180 185 190

Val Asp Val Asp Asp Leu His Ile Ile Val Gln Glu Leu Arg Gly Ser
195 200 205

Ile Leu Asp Ala Met Arg Pro
210 215

<210> 86
<211> 831
<212> PRT
<213> Homo sapiens

<400> 86
Met Lys Val His Met His Thr Lys Phe Cys Leu Ile Cys Leu Leu Thr
1 5 10 15

Phe Ile Phe His His Cys Asn His Cys His Glu Glu His Asp His Gly
20 25 30

Pro Glu Ala Leu His Arg Gln His Arg Gly Met Thr Glu Leu Glu Pro
35 40 45

Ser Lys Phe Ser Lys Gln Ala Ala Glu Asn Glu Lys Lys Tyr Tyr Ile
50 55 60

Glu Lys Leu Phe Glu Arg Tyr Gly Glu Asn Gly Arg Leu Ser Phe Phe
65 70 75 80

Gly Leu Glu Lys Leu Leu Thr Asn Leu Gly Leu Gly Glu Arg Lys Val
85 90 95

Val Glu Ile Asn His Glu Asp Leu Gly His Asp His Val Ser His Leu
100 105 110

Asp Ile Leu Ala Val Gln Glu Gly Lys His Phe His Ser His Asn His
115 120 125

Gln His Ser His Asn His Leu Asn Ser Glu Asn Gln Thr Val Thr Ser
130 135 140

Val Ser Thr Lys Arg Asn His Lys Cys Asp Pro Glu Lys Glu Thr Val
145 150 155 160

Glu Val Ser Val Lys Ser Asp Asp Lys His Met His Asp His Asn His
165 170 175

Arg Leu Arg His His His Arg Leu His His His Leu Asp His Asn Asn
180 185 190

Thr His His Phe His Asn Asp Ser Ile Thr Pro Ser Glu Arg Gly Glu
195 200 205

Pro Ser Asn Glu Pro Ser Thr Glu Thr Asn Lys Thr Gln Glu Gln Ser
210 215 220

Asp Val Lys Leu Pro Lys Gly Lys Arg Lys Lys Lys Gly Arg Lys Ser
225 230 235 240

Asn Glu Asn Ser Glu Val Ile Thr Pro Gly Phe Pro Pro Asn His Asp
245 250 255

Gln Gly Glu Gln Tyr Glu His Asn Arg Val His Lys Pro Asp Arg Val
260 265 270

His Asn Pro Gly His Ser His Val His Leu Pro Glu Arg Asn Gly His